# Relaxation Method
# For Calculating
# Quantum Entanglement

Robert R. Tucci

P.O. Box 226

Bedford, MA 01730

tucci@ar-tiste.com

February 1, 2008

## Abstract

In a previous paper, we showed how entanglement of formation can be defined as a minimum of the quantum conditional mutual information (a.k.a. quantum conditional information transmission). In classical information theory, the Arimoto-Blahut method is one of the preferred methods for calculating extrema of mutual information. In this paper, we present a new method, akin to the Arimoto-Blahut method, for calculating entanglement of formation. We also present several examples computed with a computer program called Causa Común that implements the ideas of this paper.

# 1  Introduction

This paper continues a series of papers[1]-[3] investigating the connection between quantum entanglement and conditional information transmission (a.k.a conditional mutual information, abbreviated CMI). In the last paper of that series, we expressed the entanglement of formation as a minimum of the quantum CMI. Eureka! In classical information theory, one of the preferred methods for numerically calculating extrema of mutual information is the Arimoto-Blahut algorithm [4]-[6]. One wonders whether something akin to that algorithm can be used to calculate entanglement. After much huffing and puffing, we have found the answer to be yes. In this paper we present our first results. More specifically, we present a new algorithm that yields the entanglement of formation of any bi-partite density matrix and a corresponding optimum decomposition of that density matrix. Generalization of the algorithm to n-partite systems appears straightforward but we do not address it here. We also describe a C++ computer program called Causa Común that implements the ideas of this paper. Finally, we present some examples computed with Causa Común.

Prior to us, as far as we know, only one group of researchers[7] has ever used a quantum version of the Arimoto-Blahut algorithm. They used it to calculate quantum channel capacities.

There exist other excellent computer programs, written prior to ours, that can calculate various features of quantum entanglement. See Ref.[8] and [9]. The software described in Ref.[9] also calculates entanglement of formation and optimal decompositions, but it uses a conjugate gradient method that is very different from ours.

# 2  Notation

In this section, we will introduce certain notation which is used throughout the paper.

Let $Bool = \{0, 1\}$. For any finite set $S$, let $|S|$ denote the number of elements in $S$. The Kronecker delta function $\delta(x, y)$ equals one if $x = y$ and zero otherwise. We will often abbreviate $\delta(x, y)$ by $\delta_y^x$. Sometimes we will replace an index by the symbol "$\bullet$". By this we mean that all values of the index are included. For example, if we are dealing with $w_\alpha$ where $\alpha \in \{1, 2, \dots n\}$, $w_\bullet$ will represent the vector $(w_1, w_2, \dots, w_n)$. For any Hilbert space $\mathcal{H}$, $dim(\mathcal{H})$ will stand for the dimension of $\mathcal{H}$. If $|\psi\rangle \in \mathcal{H}$, then we will often represent the projection operator $|\psi\rangle\langle\psi|$ by $\pi(\psi)$.

We will underline random variables. For example, we might write $P(\underline{x} = x)$ or $P_{\underline{x}}(x)$ for the probability that the random variable $\underline{x}$ assumes value $x$. $P(\underline{x} = x) = P_{\underline{x}}(x)$ will often be abbreviated by $P(x)$ when no confusion is likely. $S_{\underline{x}}$ will denote the set of values which the random variable $\underline{x}$ may assume, and $N_{\underline{x}}$ will denote the number of elements in $S_{\underline{x}}$. With each random variable $\underline{x}$, we will associate an orthonormal basis $\{|x\rangle | x \in S_{\underline{x}}\}$ which we will call the $\underline{x}$ *basis*. $\mathcal{H}_{\underline{x}}$ will represent the Hilbert space spanned by the $\underline{x}$ basis. For any $|\psi_{\underline{x}}\rangle \in \mathcal{H}_{\underline{x}}$, we will use $\psi_x$ to represent

$\langle x|\psi_{\underline{x}}\rangle$.

For any two random variables $\underline{x}$ and $\underline{y}$, $S_{\underline{x},\underline{y}}$ will represent the direct product set $S_{\underline{x}} \times S_{\underline{y}} = \{(x,y)|x \in S_{\underline{x}}, y \in S_{\underline{y}}\}$. Furthermore, $\mathcal{H}_{\underline{x},\underline{y}}$ will represent $\mathcal{H}_{\underline{x}} \otimes \mathcal{H}_{\underline{y}}$, the tensor product of Hilbert spaces $\mathcal{H}_{\underline{x}}$ and $\mathcal{H}_{\underline{y}}$. If $|x\rangle$ for all $x$ is the $\underline{x}$ basis and $|y\rangle$ for all $y$ is the $\underline{y}$ basis, then $\mathcal{H}_{\underline{xy}}$ is the vector space spanned by $\{|x,y\rangle|x \in S_{\underline{x}}, y \in S_{\underline{y}}\}$, where $|x,y\rangle = |x\rangle|y\rangle$.

pd($S_{\underline{x}}$) will denote the set of all probability distributions $P_{\underline{x}}$ for the random variable $\underline{x}$; i.e., all functions $P_{\underline{x}} : S_{\underline{x}} \to [0,1]$ such that $\sum_x P_{\underline{x}}(x) = 1$. dm($\mathcal{H}_{\underline{x}}$) will denote the set of all density matrices $\rho_{\underline{x}}$ acting on the Hilbert space $\mathcal{H}_{\underline{x}}$; i.e., the set of all $N_{\underline{x}}$ dimensional Hermitian matrices with unit trace and non-negative eigenvalues.

Whenever we use the word "ditto", as in "X (ditto, Y)", we mean that the statement is also true if X is replaced by Y. For example, if we say "A (ditto, X) is smaller than B (ditto, Y)", we mean "A is smaller than B" and "X is smaller than Y".

Given any function $f(x)$ defined for all $x \in A$, we define

$$\frac{f(x)}{\sum_{x \in A} numerator} = \frac{f(x)}{\sum_{x \in A} f(x)} \ . \tag{1}$$

This is just a shorthand, useful when $f(x)$ is a long expression, to avoid writing $f(x)$ explicitly twice.

This paper will also utilize certain notation associated with classical and quantum entropy. See Refs.[10],[11] for definitions and examples of such notation. In particular, we will assume that the reader is familiar with the definition of the classical entropies $H(\underline{x})$, $H(\underline{x}|\underline{y})$ (conditional entropy) and $H(\underline{x}:\underline{y})$ (mutual entropy) associated with any $P_{\underline{xy}} \in$ pd($S_{\underline{xy}}$). We will also assume that the reader is familiar with the definitions of the quantum entropies $S_{\rho_{\underline{xy}}}(\underline{x})$, $S_{\rho_{\underline{xy}}}(\underline{x}|\underline{y})$ and $S_{\rho_{\underline{xy}}}(\underline{x}:\underline{y})$ associated with any $\rho_{\underline{xy}} \in$ dm($\mathcal{H}_{\underline{xy}}$).

For $P_{\underline{x}}, P'_{\underline{x}} \in$ pd($S_{\underline{x}}$), the *classical Kullback-Leibler (KL) distance* is defined by

$$D(P_{\underline{x}}//P'_{\underline{x}}) = \sum_x P(x) \log_2[P(x)/P'(x)] \ . \tag{2}$$

For $\rho_{\underline{x}}, \rho'_{\underline{x}} \in$ dm($\mathcal{H}_{\underline{x}}$), the *quantum KL distance* is defined by

$$D(\rho_{\underline{x}}//\rho'_{\underline{x}}) = \text{tr}_{\underline{x}}[\rho_{\underline{x}}(\log_2 \rho_{\underline{x}} - \log_2 \rho'_{\underline{x}})] \ . \tag{3}$$

The classical (and quantum) KL distance is always non-negative and equals zero iff its two arguments are equal. It has many other useful properties. For more information about the KL distance, see [11] for the classical case and [12] for the quantum one.

When discussing classical physics (ditto, quantum physics), we will refer to various probability distributions (ditto, density matrices) which are "descendants" of (i.e., can be derived from) a parent $P_{\underline{xy\alpha}} \in$ pd($S_{\underline{xy\alpha}}$) (ditto, $\rho^\alpha_{\underline{xy}} \in$ dm($\mathcal{H}_{\underline{xy}}$)). As an aid to the reader, here is a table mapping the classical descendants to their quantum

counterparts. The reader may find it helpful to continue returning to this "Cast of Characters" table as he advances through this play.

| Classical | Quantum |
|---|---|
| $\{P(x,y|\alpha)|\forall x,y\} \in \mathrm{pd}(S_{\underline{xy}})$ | $\rho_{\underline{xy}}^{\alpha} \in \mathrm{dm}(\mathcal{H}_{\underline{xy}})$ |
| $\{P(\alpha)|\forall\alpha\} \in \mathrm{pd}(S_{\underline{\alpha}})$ | $\{w_{\alpha}|\forall\alpha\} \in \mathrm{pd}(S_{\underline{\alpha}})$ |
| $P(x,y,\alpha) = P(x,y|\alpha)P(\alpha)$ | $K_{\underline{xy}}^{\alpha} = w_{\alpha}\rho_{\underline{xy}}^{\alpha}$ |
| | $\rho_{\underline{xy\alpha}} = \sum_{\alpha}|\alpha\rangle\langle\alpha|K_{\underline{xy}}^{\alpha}$ |
| $P(x|\alpha) = \sum_{y} P(x,y|\alpha)$ | $\rho_{\underline{x}}^{\alpha} = \mathrm{tr}_{\underline{y}}\rho_{\underline{xy}}^{\alpha}$ |
| $P(x,\alpha) = \sum_{y} P(x,y,\alpha)$ | $K_{\underline{x}}^{\alpha} = \mathrm{tr}_{\underline{y}}K_{\underline{xy}}^{\alpha}$ |
| $P(y|\alpha) = \sum_{x} P(x,y|\alpha)$ | $\rho_{\underline{y}}^{\alpha} = \mathrm{tr}_{\underline{x}}\rho_{\underline{xy}}^{\alpha}$ |
| $P(y,\alpha) = \sum_{x} P(x,y,\alpha)$ | $K_{\underline{y}}^{\alpha} = \mathrm{tr}_{\underline{x}}K_{\underline{xy}}^{\alpha}$ |
| $P(x,y) = \sum_{\alpha} P(x,y,\alpha)$ | $\rho_{\underline{xy}} = \sum_{\alpha} K_{\underline{xy}}^{\alpha}$ |
| $R(x,y,\alpha) = \frac{P(x,\alpha)P(y,\alpha)}{P(\alpha)}$ | $R_{\underline{xy}}^{\alpha} = \frac{K_{\underline{x}}^{\alpha}K_{\underline{y}}^{\alpha}}{w_{\alpha}}$ |

In the classical case, we will use a functional $R[P_{\underline{xy\alpha}}]$ of $P_{\underline{xy\alpha}}$ defined by

$$R[P_{\underline{xy\alpha}}](x,y,\alpha) = \frac{P(x,\alpha)P(y,\alpha)}{P(\alpha)} = P(x|\alpha)P(y|\alpha)P(\alpha) . \tag{4}$$

We will often abbreviate $R[P_{\underline{xy\alpha}}]$ by $R$ if no confusion is likely. Note that $R(x,y,\alpha) \geq 0$ and $\sum_{x,y,\alpha} R(x,y,\alpha) = 1$ so $R \in \mathrm{pd}(S_{\underline{xy\alpha}})$. In the quantum case, we will use a functional $R_{\underline{xy}}^{\alpha}[K_{\underline{xy}}^{\alpha}]$ of $K_{\underline{xy}}^{\alpha}$ defined by

$$R_{\underline{xy}}^{\alpha}[K_{\underline{xy}}^{\alpha}] = \frac{K_{\underline{x}}^{\alpha}K_{\underline{y}}^{\alpha}}{w_{\alpha}} = w_{\alpha}\rho_{\underline{x}}^{\alpha}\rho_{\underline{y}}^{\alpha} . \tag{5}$$

We will often abbreviate $R_{\underline{xy}}^{\alpha}[K_{\underline{xy}}^{\alpha}]$ by $R_{\underline{xy}}^{\alpha}$ if no confusion is likely. Note that $R_{\underline{xy}}^{\alpha}/w_{\alpha} \in \mathrm{dm}(\mathcal{H}_{\underline{xy}})$ for all $\alpha$.

Suppose $\rho_{\underline{xy}} \in \mathrm{dm}(\mathcal{H}_{\underline{xy}})$ has eigensystem $\{(\lambda_{j}, |\phi_{j}\rangle)|\forall j\}$. Thus,

$$\rho_{\underline{xy}} = \sum_{j} \lambda_{j}|\phi_{j}\rangle\langle\phi_{j}| . \tag{6}$$

According to Ref.[17], $\rho_{\underline{xy}}$ can be expressed as

$$\rho_{\underline{xy}} = \sum_{\alpha} w_{\alpha}|\psi_{\alpha}\rangle\langle\psi_{\alpha}| , \tag{7}$$

where $w_{\bullet} \in \mathrm{pd}(S_{\underline{\alpha}})$, and $|\psi_{\alpha}\rangle \in \mathcal{H}_{\underline{xy}}$ for all $\alpha$, if and only if there exists a transformation $T_{j}^{\alpha}$ ($\alpha \in S_{\underline{\alpha}}$, $j \in S_{\underline{xy}}$) which is "right unitary":

4

$$\sum_{\alpha} T_j^{\alpha} T_{j'}^{\alpha *} = \delta_j^{j'} \ , \tag{8}$$

and which satisfies

$$\sum_{j} T_j^{\alpha} \sqrt{\lambda_j} |\phi_j\rangle = \sqrt{w_{\alpha}} |\psi_{\alpha}\rangle \ . \tag{9}$$

Suppose $A : \mathcal{H} \to \mathcal{H}$ is an operator with eigensystem $\{(\lambda_j, |\phi_j\rangle) | \forall j\}$. Thus,

$$A = \sum_{j} \lambda_j |\phi_j\rangle\langle\phi_j| \ . \tag{10}$$

The support (ditto, kernel) of $A$ is the subspace of $\mathcal{H}$ consisting of the zero vector and all those vectors in $\mathcal{H}$ for which $A$ does not (ditto, does) vanish. Suppose $\epsilon$ is a very small positive number and $\chi_{[0,\epsilon]}(x)$ for real $x$ is an indicator function that equals 1 if $x \in [0, \epsilon]$ and vanishes otherwise. Then we define the projectors $\pi_{ker}(A)$ and $\pi_{supp}(A)$ by

$$\pi_{ker}(A) = \sum_{j} \chi_{[0,\epsilon]}(\lambda_j) |\phi_j\rangle\langle\phi_j| \ , \tag{11}$$

$$\pi_{supp}(A) = 1 - \pi_{ker}(A) \ . \tag{12}$$

# 3   Classical Physics Minimization

In this section, we will discuss a minimization of the CMI for classical probabilities.

The CMI for $P_{\underline{xy\alpha}}$ can be expressed as

$$
\begin{aligned}
H(\underline{x} : \underline{y} | \underline{\alpha}) &= \sum_{x,y,\alpha} P(x, y, \alpha) \log_2 \left( \frac{P(x, y|\alpha)}{P(x|\alpha)P(y|\alpha)} \right) \\
&= \sum_{x,y,\alpha} P(x, y, \alpha) \log_2 \left( \frac{P(x, y, \alpha)P(\alpha)}{P(x, \alpha)P(y, \alpha)} \right) \\
&= \sum_{x,y,\alpha} P(x, y, \alpha) \log_2 \left( \frac{P(x, y, \alpha)}{R(x, y, \alpha)} \right) \ .
\end{aligned}
\tag{13}
$$

Let $\mathcal{P}_{cla}$ be the set of those $P_{\underline{xy\alpha}} \in \mathrm{pd}(S_{\underline{xy\alpha}})$ for which the sum over $\alpha$ of $P(x, y, \alpha)$ equals a fixed $\tilde{P}_{\underline{xy}} \in \mathrm{pd}(S_{\underline{xy}})$:

$$\mathcal{P}_{cla} = \{ P_{\underline{xy\alpha}} \in \mathrm{pd}(S_{\underline{xy\alpha}}) | P_{\underline{xy}} = \tilde{P}_{\underline{xy}} \} \ . \tag{14}$$

We define the entanglement $E_{cla}$ by

$$E_{cla} = \left( \frac{1}{2} \right) \min_{P_{\underline{xy\alpha}} \in \mathcal{P}_{cla}} H(\underline{x} : \underline{y} | \underline{\alpha}) \ . \tag{15}$$

In this definition of $E_{cla}$, $N_{\underline{\alpha}}$ is assumed to be fixed . Clearly, for $N_{\underline{\alpha}}$ large enough, $E_{cla} = 0$. Indeed, suppose $N_{\underline{\alpha}} = N_{\underline{x}} N_{\underline{y}}$ and $\alpha \to (x_\alpha, y_\alpha)$ is a 1-1 onto function from $S_{\underline{\alpha}}$ to $S_{\underline{xy}}$. Then $H(\underline{x} : \underline{y} | \underline{\alpha}) = 0$ for $P(x, y, \alpha) = \tilde{P}(x_\alpha, y_\alpha) \delta_x^{x_\alpha} \delta_y^{y_\alpha}$.

It is convenient to consider the following "Lagrangian" functional of two probability distributions:

$$\mathcal{L}(P_{\underline{xy\alpha}}, P'_{\underline{xy\alpha}}) = \sum_{x,y,\alpha} P(x, y, \alpha) \ln \left( \frac{P(x, y, \alpha)}{R'(x, y, \alpha)} \right) , \tag{16}$$

where $R' = R[P'_{\underline{xy\alpha}}]$. $E_{cla}$ can be defined in terms of this Lagrangian by

$$E_{cla} = \left( \frac{1}{2 \ln 2} \right) \min_{P_{\underline{xy\alpha}} \in \mathcal{P}_{cla}} \mathcal{L}(P_{\underline{xy\alpha}}, P_{\underline{xy\alpha}}) . \tag{17}$$

**Lemma 3.1** $\mathcal{L}(P_{\underline{xy\alpha}}, P'_{\underline{xy\alpha}})$ *is convex ($\cup$) in its first argument.*

**proof:**

The proof is very similar to the proof that the classical entropy $H(P)$ is concave ($\cap$) in $P$. Let $f(x) = x \ln x$. If we can show that $\Delta \mathcal{L}(P) = \sum_{x,y,\alpha} f(P(x, y, \alpha))$ is convex in $P$, we will be done, because the remaining part $\mathcal{L} - \Delta \mathcal{L}$ is linear in $P$. Let $\lambda \in [0, 1]$, $\bar{\lambda} = 1 - \lambda$, and $P = \lambda P^{(1)} + \bar{\lambda} P^{(2)}$, where $P^{(1)}, P^{(2)} \in pd(S_{\underline{xy\alpha}})$. Since $f(x)$ is convex in $x$,

$$
\begin{aligned}
\Delta \mathcal{L}(P) &= \sum_{x,y,\alpha} f(P(x, y, \alpha)) \\
&\leq \lambda \sum_{x,y,\alpha} f\left(P^{(1)}(x, y, \alpha)\right) + \bar{\lambda} \sum_{x,y,\alpha} f\left(P^{(2)}(x, y, \alpha)\right) \\
&= \lambda \Delta \mathcal{L}(P^{(1)}) + \bar{\lambda} \Delta \mathcal{L}(P^{(2)}) .
\end{aligned}
\tag{18}
$$

QED

(On the other hand, $\mathcal{L}(P, P')$ is not generally convex or concave in its second argument. This can be seen by taking the second derivative of $\mathcal{L}$ with respect to that argument.)

**Theorem 3.1** *Let*

$$\mathcal{P}'_{cla} = \mathrm{pd}(S_{\underline{xy\alpha}}) . \tag{19}$$

*At fixed $P_{\underline{xy\alpha}} \in \mathcal{P}_{cla}$, $\mathcal{L}(P_{\underline{xy\alpha}}, P'_{\underline{xy\alpha}})$ is minimized over all $P'_{\underline{xy\alpha}} \in \mathcal{P}'_{cla}$ iff $P'_{\underline{xy\alpha}}$ satisfies*

$$P'_{\underline{\alpha}} = P_{\underline{\alpha}} , \tag{20a}$$

$$P'_{\underline{x\alpha}} = P_{\underline{x\alpha}} , \tag{20b}$$

*and*

$$P'_{\underline{y\alpha}} = P_{\underline{y\alpha}} \ . \tag{20c}$$

*Thus,*

$$\mathcal{L}(P_{\underline{xy\alpha}}, P_{\underline{xy\alpha}}) = \min_{P'_{\underline{xy\alpha}} \in \mathcal{P}'_{cla}} \mathcal{L}(P_{\underline{xy\alpha}}, P'_{\underline{xy\alpha}}) \ , \tag{21}$$

*and*

$$E_{cla} = \left(\frac{1}{2\ln 2}\right) \min_{P_{\underline{xy\alpha}} \in \mathcal{P}_{cla}} \min_{P'_{\underline{xy\alpha}} \in \mathcal{P}'_{cla}} \mathcal{L}(P_{\underline{xy\alpha}}, P'_{\underline{xy\alpha}}) \ . \tag{22}$$

**proof:**

The Lagrangian $\mathcal{L}$ can be expressed in terms of the KL distance as follows:

$$\frac{1}{\ln 2}\mathcal{L}(P_{\underline{xy\alpha}}, P'_{\underline{xy\alpha}}) = \begin{cases} D(P_{\underline{\alpha}}//P'_{\underline{\alpha}}) \\ + \sum_\alpha P(\alpha)D\left(P(x,y|\alpha)//P(x|\alpha)P(y|\alpha)\right) \\ + \sum_\alpha P(\alpha)D\left(P(x|\alpha)//P'(x|\alpha)\right) \\ + \sum_\alpha P(\alpha)D\left(P(y|\alpha)//P'(y|\alpha)\right) \end{cases} \ . \tag{23}$$

The KL distance is always non-negative and equals zero iff its two arguments are equal. Hence, Eqs.(20) are necessary and sufficient conditions for $\mathcal{L}(P, P')$ to have a global minimum in $P'$ at fixed $P$. QED

**Theorem 3.2** *At fixed $P'_{\underline{xy\alpha}} \in \mathcal{P}'{}_{cla}$, $\mathcal{L}(P_{\underline{xy\alpha}}, P'_{\underline{xy\alpha}})$ is minimized over all $P_{\underline{xy\alpha}} \in \mathcal{P}_{cla}$ iff $P_{\underline{xy\alpha}}$ satisfies*

$$P(\alpha|x,y) = R'(\alpha|x,y) \tag{24}$$

*for all $x, y, \alpha$.*

**proof:**

Suppose a minimum is achieved. Define a new Lagrangian $\mathcal{L}_{tot}$ by adding to $\mathcal{L}$ a Lagrange multiplier term that enforces the constraint that the sum over $\alpha$ of $P(x, y, \alpha)$ equals a fixed probability distribution $\tilde{P}(x, y) \in \text{pd}(S_{\underline{xy}})$:

$$\mathcal{L}_{tot} = \mathcal{L}(P_{\underline{xy\alpha}}, P'_{\underline{xy\alpha}}) + \sum_{x,y} \lambda(x,y)[\sum_\alpha P(x,y,\alpha) - \tilde{P}(x,y)] \ . \tag{25}$$

$\mathcal{L}_{tot}$ should not change if we vary infinitesimally and independently the quantities $\lambda(x, y)$ and $P(x, y, \alpha)$ for all $x, y, \alpha$. Thus,

$$0 = \frac{\partial \mathcal{L}_{tot}}{\partial P(x_o, y_o, \alpha_o)} = \ln P(x_o, y_o, \alpha_o) + 1 - \ln R'(x_o, y_o, \alpha_o) + \lambda(x_o, y_o) \ . \tag{26}$$

Let

$$\Delta(x, y) = -1 - \lambda(x, y) . \tag{27}$$

Then

$$\ln P(x, y, \alpha) = \ln R'(x, y, \alpha) + \Delta(x, y) . \tag{28}$$

Taking the exponential of both sides of the last equation and summing them over $\alpha$ yields the following constraint on $\Delta(x, y)$:

$$P(x, y) = \sum_\alpha \exp[\ln R'(x, y, \alpha) + \Delta(x, y)] . \tag{29}$$

Solving the last equation for $\Delta(x, y)$ yields:

$$\Delta(x, y) = -\ln\left(\frac{R'(x, y)}{P(x, y)}\right) . \tag{30}$$

Eq.(24) now follows from Eqs.(28) and (30).

$\mathcal{L}(P, P')$ has an extremum in $P$ at fixed $P'$ iff Eq.(24) is true. Furthermore, since $\mathcal{L}$ is convex in its first argument, the extremum must be a global minimum. QED

**Theorem 3.3** *The CMI minimum which defines $E_{cla}$ is achieved iff*

$$P(x, y, \alpha) = P(x, y)\frac{R(x, y, \alpha)}{\sum_\alpha R(x, y, \alpha)} . \tag{31}$$

*Furthermore,*

$$E_{cla} = \left(\frac{1}{2\ln 2}\right)\langle\Delta\rangle , \tag{32}$$

*where*

$$\Delta(x, y) = -\ln\left(\frac{R(x, y)}{P(x, y)}\right) , \tag{33}$$

*and*

$$\langle\Delta\rangle = \sum_{x,y} P(x, y)\Delta(x, y) . \tag{34}$$

*This justifies calling $\Delta$ an "entanglement operator".*

**proof:**

The first part of this claim just brings together results obtained in the previous two theorems. The second part where $E_{cla}$ is expressed in terms of $\Delta$ follows from:

$$
\begin{aligned}
E_{cla} &= \left(\frac{1}{2\ln 2}\right) \sum_{x,y,\alpha} P(x,y,\alpha) \ln \frac{P(x,y,\alpha)}{R(x,y,\alpha)} \\
&= \left(\frac{1}{2\ln 2}\right) \sum_{x,y,\alpha} P(x,y,\alpha)\Delta(x,y) .
\end{aligned}
\tag{35}
$$

QED

The last theorem gives certain conditions obeyed by any $P_{\underline{xy\alpha}}$ which achieves $E_{cla}$. Next we will define a sequence of probability distributions, $P_{\underline{xy\alpha}}^{(n)}$ for $n = 0, 1, \ldots$. The sequence will converge to $P_{\underline{xy\alpha}}$ as $n \to \infty$. We will define our sequence recursively. In the following diagram, each quantity is defined in terms of the quantities that point to it.

$$
P_{\underline{xy\alpha}}^{(0)} \to P_{\underline{xy\alpha}}^{(1)} \to P_{\underline{xy\alpha}}^{(2)} \to \cdots
\tag{36}
$$

Let $P_{\underline{xy\alpha}}^{(0)}$ be chosen arbitrarily from $\mathcal{P}_{cla}$. For any $n \geq 0$, let

$$
P^{(n+1)}(x,y,\alpha) = P(x,y) \left[ \frac{R^{(n)}(x,y,\alpha)}{\sum_\alpha R^{(n)}(x,y,\alpha)} \right] ,
\tag{37}
$$

where $R^{(n)} = R[P_{\underline{xy\alpha}}^{(n)}]$.

In this paper, we won't prove that the sequence of $P_{\underline{xy\alpha}}^{(n)}$ converges. We defer that to future papers, confining ourselves here to presenting some empirical and intuitive motivations for the sequence. In Section 6, we give some computer results that are good empirical evidence of convergence. Note that if the limit of the sequence does exist, then the limit of Eq.(37) is Eq.(31).

# 4    Quantum Physics, Mixed Minimization

In this section, we will discuss a quantum counterpart of the classical minimization problem discussed in the previous section.

Consider all $\rho_{\underline{xy\alpha}} \in \mathrm{dm}(\mathcal{H}_{\underline{xy\alpha}})$ of the special form:

$$
\rho_{\underline{xy\alpha}} = \sum_\alpha |\alpha\rangle\langle\alpha| w_\alpha \rho_{\underline{xy}}^\alpha = \sum_\alpha |\alpha\rangle\langle\alpha| K_{\underline{xy}}^\alpha ,
\tag{38}
$$

where $|\alpha\rangle$ for all $\alpha$ is an orthonormal basis of $\mathcal{H}_{\underline{\alpha}}$, $w_\bullet \in \mathrm{pd}(S_\alpha)$, and $\rho_{\underline{xy}}^\alpha \in \mathrm{dm}(\mathcal{H}_{\underline{xy}})$. As shown in Ref.[3], the CMI for $\rho_{\underline{xy\alpha}}$ can be expressed as

$$
\begin{aligned}
S_{\rho_{\underline{xy\alpha}}}(\underline{x}:\underline{y}|\underline{\alpha}) &= \sum_\alpha w_\alpha S_{\rho_{\underline{xy}}^\alpha}(\underline{x}:\underline{y}) \\
&= \sum_\alpha w_\alpha [S(\rho_{\underline{x}}^\alpha) + S(\rho_{\underline{y}}^\alpha) - S(\rho_{\underline{xy}}^\alpha)]
\end{aligned}
$$

9

$$= \sum_{\alpha} \text{tr}_{\underline{x},\underline{y}} \left[ K_{\underline{xy}}^{\alpha} (\log_2 K_{\underline{xy}}^{\alpha} - \log_2 \frac{K_{\underline{x}}^{\alpha} K_{\underline{y}}^{\alpha}}{w_{\alpha}}) \right]$$

$$= \sum_{\alpha} \text{tr}_{\underline{x},\underline{y}} \left[ K_{\underline{xy}}^{\alpha} (\log_2 K_{\underline{xy}}^{\alpha} - \log_2 R_{\underline{xy}}^{\alpha}) \right] . \tag{39}$$

Let

$$\mathcal{K}_{mixed}^{\star} = \{ K_{\underline{xy}}^{\bullet} | \forall \alpha, K_{\underline{xy}}^{\alpha} = w_{\alpha} \rho_{\underline{xy}}^{\alpha}, w_{\bullet} \in \text{pd}(S_{\underline{\alpha}}), \rho_{\underline{xy}}^{\alpha} \in \text{dm}(\mathcal{H}_{\underline{xy}}) \} , \tag{40}$$

and

$$\mathcal{K}_{mixed} = \{ K_{\underline{xy}}^{\bullet} \in \mathcal{K}_{mixed}^{\star} | \sum_{\alpha} K_{\underline{xy}}^{\alpha} = \tilde{\rho}_{\underline{xy}} \} . \tag{41}$$

We define the entanglement $E_{mixed}$ by

$$E_{mixed} = \left( \frac{1}{2} \right) \min_{K_{\underline{xy}}^{\bullet} \in \mathcal{K}_{mixed}} S_{\rho_{\underline{xy\alpha}}}(\underline{x} : \underline{y} | \underline{\alpha}) . \tag{42}$$

In this definition of $E_{mixed}$, $N_{\underline{\alpha}}$ will be assumed to tend to infinity. Ref.[14] shows that the limit is reached at a finite $N_{\underline{\alpha}} \leq (N_{\underline{x}} N_{\underline{y}})^2$.

It is convenient to consider the following "Lagrangian" functional of two density matrices:

$$\mathcal{L}(K_{\underline{xy}}^{\alpha}, K_{\underline{xy}}^{'\alpha}) = \sum_{\alpha} \text{tr}_{\underline{x},y} \left[ K_{\underline{xy}}^{\alpha} (\ln K_{\underline{xy}}^{\alpha} - \ln R_{\underline{xy}}^{'\alpha}) \right] , \tag{43}$$

where $R_{\underline{xy}}^{'\alpha} = R_{\underline{xy}}^{\alpha}[K_{\underline{xy}}^{'\alpha}]$. $E_{mixed}$ can be defined in terms of this Lagrangian by

$$E_{mixed} = \left( \frac{1}{2 \ln 2} \right) \min_{K_{\underline{xy}}^{\bullet} \in \mathcal{K}_{mixed}} \mathcal{L}(K_{\underline{xy}}^{\alpha}, K_{\underline{xy}}^{\alpha}) . \tag{44}$$

**Lemma 4.1** $\mathcal{L}(K_{\underline{xy}}^{\alpha}, K_{\underline{xy}}^{'\alpha})$ *is convex ($\cup$) in its first argument.*

**proof:**

The proof is very similar to the proof that the quantum entropy $S(\rho)$ is concave ($\cap$) in $\rho$ (see Ref.[12]). Let $f(x) = x \ln x$. If we can show that $\Delta \mathcal{L}(K_{\underline{xy}}^{\alpha}) = \sum_{\alpha} \text{tr}_{\underline{xy}} f(K_{\underline{xy}}^{\alpha})$ is convex in $K_{\underline{xy}}^{\alpha}$, we will be done, because the remaining part $\mathcal{L} - \Delta \mathcal{L}$ is linear in $K_{\underline{xy}}^{\alpha}$. Let $\lambda \in [0, 1]$, $\bar{\lambda} = 1 - \lambda$, and $K^{\alpha} = \lambda K^{(1)\alpha} + \bar{\lambda} K^{(2)\alpha}$. Here $K^{(1)\alpha}$ and $K^{(2)\alpha}$ belong to $\text{dm}(\mathcal{H}_{\underline{xy}})$ if we normalize them by dividing them by their trace. Let $K^{\alpha}$ have eigensystem $\{(m_j^{\alpha}, |\phi_j^{\alpha}\rangle) | \forall j\}$. Since $f(x)$ is convex in $x$,

$$\begin{aligned}
\Delta\mathcal{L}(K^\alpha) &= \sum_{\alpha,j} f(m_j^\alpha) = \sum_{\alpha,j} f(\langle\phi_j^\alpha|K^\alpha|\phi_j^\alpha\rangle) \\
&\leq \lambda \sum_{\alpha,j} f(\langle\phi_j^\alpha|K^{(1)\alpha}|\phi_j^\alpha\rangle) + \bar\lambda \sum_{\alpha,j} f(\langle\phi_j^\alpha|K^{(2)\alpha}|\phi_j^\alpha\rangle) \\
&\leq \lambda \sum_{\alpha,j} \langle\phi_j^\alpha|f(K^{(1)\alpha})|\phi_j^\alpha\rangle + \bar\lambda \sum_{\alpha,j} \langle\phi_j^\alpha|f(K^{(2)\alpha})|\phi_j^\alpha\rangle \\
&= \lambda \Delta\mathcal{L}(K^{(1)\alpha}) + \bar\lambda \Delta\mathcal{L}(K^{(2)\alpha}) \; .
\end{aligned} \tag{45}$$

QED

**Theorem 4.1** *Let*

$$\mathcal{K}'_{mixed} = \mathcal{K}^\star_{mixed} \; . \tag{46}$$

*At fixed $K_{\underline{xy}}^\bullet \in \mathcal{K}_{mixed}$, $\mathcal{L}(K_{\underline{xy}}^\alpha, K_{\underline{xy}}'^\alpha)$ is minimized over all $K_{\underline{xy}}'^\bullet \in \mathcal{K}'_{mixed}$ iff $K_{\underline{xy}}'^\bullet$ satisfies*

$$w'_\alpha = w_\alpha \; , \tag{47a}$$

$$K_{\underline{x}}'^\alpha = K_{\underline{x}}^\alpha \; , \tag{47b}$$

*and*

$$K_{\underline{y}}'^\alpha = K_{\underline{y}}^\alpha \; . \tag{47c}$$

*Thus,*

$$\mathcal{L}(K_{\underline{xy}}^\alpha, K_{\underline{xy}}^\alpha) = \min_{K_{\underline{xy}}'^\bullet \in \mathcal{K}'_{mixed}} \mathcal{L}(K_{\underline{xy}}^\alpha, K_{\underline{xy}}'^\alpha) \; , \tag{48}$$

*and*

$$E_{mixed} = \left(\frac{1}{2\ln 2}\right) \min_{K_{\underline{xy}}^\bullet \in \mathcal{K}_{mixed}} \; \min_{K_{\underline{xy}}'^\bullet \in \mathcal{K}'_{mixed}} \mathcal{L}(K_{\underline{xy}}^\alpha, K_{\underline{xy}}'^\alpha) \; . \tag{49}$$

**proof:**
     The Lagrangian $\mathcal{L}$ can be expressed in terms of the KL distance as follows:

$$\frac{1}{\ln 2}\mathcal{L}(K_{\underline{xy}}^\alpha, K_{\underline{xy}}'^\alpha) = \begin{cases} D(w_\alpha//w'_\alpha) \\ + \sum_\alpha w_\alpha D(\rho_{\underline{xy}}^\alpha//\rho_{\underline{x}}^\alpha \rho_{\underline{y}}^\alpha) \\ + \sum_\alpha w_\alpha D(\rho_{\underline{x}}^\alpha//\rho_{\underline{x}}'^\alpha) \\ + \sum_\alpha w_\alpha D(\rho_{\underline{y}}^\alpha//\rho_{\underline{y}}'^\alpha) \end{cases} \; . \tag{50}$$

Hence, Eqs.(47) are necessary and sufficient conditions for $\mathcal{L}(K, K')$ to have a global minimum in $K'$ at fixed $K$. QED

**Theorem 4.2** *At fixed* $K'^{\bullet}_{\underline{xy}} \in \mathcal{K}'_{mixed}$, $\mathcal{L}(K^{\alpha}_{\underline{xy}}, K'^{\alpha}_{\underline{xy}})$ *is minimized over all* $K^{\bullet}_{\underline{xy}} \in \mathcal{K}_{mixed}$ *iff* $K^{\alpha}_{\underline{xy}}$ *satisfies*

$$\ln K^{\alpha}_{\underline{xy}} = \ln R'^{\alpha}_{\underline{xy}} + \Delta_{\underline{xy}} \, , \tag{51a}$$

*and*

$$\rho_{\underline{xy}} = \sum_{\alpha} \exp(\ln R'^{\alpha}_{\underline{xy}} + \Delta_{\underline{xy}}) \, . \tag{51b}$$

**proof:**
        Suppose a minimum is achieved. Define a new Lagrangian $\mathcal{L}_{tot}$ by adding to $\mathcal{L}$ a Lagrange multiplier term that enforces the constraint that the sum over $\alpha$ of $K^{\alpha}_{\underline{xy}}$ equals a fixed density matrix $\tilde{\rho}_{\underline{xy}} \in \mathrm{dm}(\mathcal{H}_{\underline{xy}})$:

$$\mathcal{L}_{tot} = \mathcal{L}(K^{\alpha}_{\underline{xy}}, K'^{\alpha}_{\underline{xy}}) + \mathrm{tr}_{\underline{x},\underline{y}}[\lambda_{\underline{xy}}(\sum_{\alpha} K^{\alpha}_{\underline{xy}} - \tilde{\rho}_{\underline{xy}})] \, . \tag{52}$$

$\mathcal{L}_{tot}$ should not change if we vary infinitesimally and independently the operators $\lambda_{\underline{xy}}$ and $K^{\alpha}_{\underline{xy}}$ for all $\alpha$. Thus,

$$0 = \delta\mathcal{L}_{tot} = \sum_{\alpha} \mathrm{tr}_{\underline{x},\underline{y}} \left[ \delta K^{\alpha}_{\underline{xy}}(\ln K^{\alpha}_{\underline{xy}} + 1 - \ln R'^{\alpha}_{\underline{xy}} + \lambda_{\underline{xy}}) \right] \, . \tag{53}$$

Let

$$\Delta_{\underline{xy}} = -1 - \lambda_{\underline{xy}} \, . \tag{54}$$

Then

$$\ln K^{\alpha}_{\underline{xy}} = \ln R'^{\alpha}_{\underline{xy}} + \Delta_{\underline{xy}} \, . \tag{55}$$

Taking the exponential of both sides of the last equation and summing them over $\alpha$ yields the following constraint on $\Delta_{\underline{xy}}$:

$$\rho_{\underline{xy}} = \sum_{\alpha} \exp[\ln R'^{\alpha}_{\underline{xy}} + \Delta_{\underline{xy}}] \, . \tag{56}$$

        $\mathcal{L}(K, K')$ has an extremum in $K$ at fixed $K'$ iff Eqs.(51) are true. Furthermore, since $\mathcal{L}$ is convex in its first argument, the extremum must be a global minimum. QED

**Theorem 4.3** *The CMI minimum which defines* $E_{mixed}$ *is achieved iff*

$$\ln K^{\alpha}_{\underline{xy}} = \ln R^{\alpha}_{\underline{xy}} + \Delta_{\underline{xy}} \, , \tag{57a}$$

*and*

$$\rho_{\underline{xy}} = \sum_\alpha \exp[\ln R^\alpha_{\underline{xy}} + \Delta_{\underline{xy}}] \ . \tag{57b}$$

*Furthermore,*

$$E_{mixed} = \left(\frac{1}{2\ln 2}\right) \langle\Delta\rangle \ , \tag{58}$$

*where*

$$\langle\Delta\rangle = \mathrm{tr}_{\underline{x},\underline{y}}(\rho_{\underline{xy}}\Delta_{\underline{xy}}) \ . \tag{59}$$

*This justifies calling $\Delta_{\underline{xy}}$ an "entanglement operator".*

**proof:**

The first part of this claim just brings together results obtained in the previous two theorems. The second part where $E_{mixed}$ is expressed in terms of $\Delta$ follows from:

$$
\begin{aligned}
E_{mixed} &= \left(\frac{1}{2\ln 2}\right) \sum_\alpha \mathrm{tr}_{\underline{x},\underline{y}}[K^\alpha_{\underline{xy}}(\ln K^\alpha_{\underline{xy}} - \ln R^\alpha_{\underline{xy}})] \\
&= \left(\frac{1}{2\ln 2}\right) \mathrm{tr}_{\underline{x},\underline{y}}(\rho_{\underline{xy}}\Delta_{\underline{xy}}) \ .
\end{aligned} \tag{60}
$$

QED

The last theorem gives certain conditions obeyed by any pair $(K^\alpha_{\underline{xy}}, \Delta_{\underline{xy}})$ which achieves $E_{mixed}$. In the classical mixed minimization problem, we were able to solve for $\Delta$ explicitly and substitute it into the remaining equations. Non-commutativity now prevents us from doing this. The way we will overcome the obstacle of non-commutativity is to solve for both $K^\alpha_{\underline{xy}}$ and $\Delta_{\underline{xy}}$ simultaneously. Next, we will define two sequences of operators, $K^{\alpha(n)}_{\underline{xy}}$ and $\Delta^{(n)}_{\underline{xy}}$ for $n = 0, 1, 2 \ldots$. The sequences will converge to $K^\alpha_{\underline{xy}}$ and $\Delta_{\underline{xy}}$, respectively, as $n \to \infty$. We will define our two sequences recursively. In the following diagram, each quantity is defined in terms of the quantities that point to it.[13]

$$
\begin{array}{ccccccc}
K^{\alpha(0)}_{\underline{xy}} & \to & K^{\alpha(1)}_{\underline{xy}} & \to & K^{\alpha(2)}_{\underline{xy}} & \cdots \\
& \nearrow & \downarrow & \nearrow & \downarrow & \\
\Delta^{(0)}_{\underline{xy}} & \to & \Delta^{(1)}_{\underline{xy}} & \to & \Delta^{(2)}_{\underline{xy}} & \cdots
\end{array} \tag{61}
$$

Roughly speaking, our strategy is: estimate $K^\alpha_{\underline{xy}}$, use the latter to get a better estimate of $\Delta_{\underline{xy}}$, use the latter to get a better estimate of $K^\alpha_{\underline{xy}}$, use the latter to get a better estimate of $\Delta_{\underline{xy}}$, and so on. Let $K^{\alpha(0)}_{\underline{xy}}$ be chosen arbitrarily from $\mathcal{K}_{mixed}$. Let $\Delta^{(0)}_{\underline{xy}} = 0$. For any $n \geq 0$, let

$$K^{\alpha(n+1)}_{\underline{xy}} = \frac{\pi_1 \exp\left[\ln R^{\alpha(n)}_{\underline{xy}} + \Delta^{(n)}_{\underline{xy}}\right] \pi_1}{\sum_\alpha \mathrm{tr}_{\underline{xy}}(numerator)} \ , \tag{62a}$$

13

(now use this $K$ to produce an even better $K$, which we call $K$ tilde:)

$$\tilde{K}_{\underline{xy}}^{\alpha(n+1)} = \frac{\pi_1 \exp\left[\ln R_{\underline{xy}}^{\alpha(n+1)} + \Delta_{\underline{xy}}^{(n)}\right] \pi_1}{\sum_\alpha \text{tr}_{\underline{xy}}(numerator)} \ , \tag{62b}$$

$$I^{(n+1)} = \left(\pi_1 \frac{1}{\sqrt{\rho_{\underline{xy}}}} \pi_1\right) \left\{\sum_\alpha \tilde{K}_{\underline{xy}}^{\alpha(n+1)}\right\} \left(\pi_1 \frac{1}{\sqrt{\rho_{\underline{xy}}}} \pi_1\right) + \pi_0 \ , \tag{62c}$$

$$\Delta_{\underline{xy}}^{(n+1)} = -\ln\left(e^{\frac{-\Delta_{\underline{xy}}^{(n)}}{2}} I^{(n+1)} e^{\frac{-\Delta_{\underline{xy}}^{(n)}}{2}}\right) \ , \tag{62d}$$

where $R_{\underline{xy}}^{\alpha(n)} = R_{\underline{xy}}^\alpha[K_{\underline{xy}}^{\alpha(n)}]$, $\pi_0 = \pi_{ker}(\rho_{\underline{xy}})$ and $\pi_1 = 1 - \pi_0$.

In this paper, we won't prove that the sequences $K_{\underline{xy}}^{\alpha(n)}$ and $\Delta_{\underline{xy}}^{(n)}$ converge. We defer that to future papers, confining ourselves here to presenting some empirical and intuitive motivations for the sequences. In Section 6, we give some computer results that are good evidence of convergence. It is easy to see that if the sequences do converge, then their limit satisfies Eqs.(57). Indeed, as $n \to \infty$, Eqs.(62) become

$$K_{\underline{xy}}^\alpha = \pi_1 \exp\left[\ln R_{\underline{xy}}^\alpha + \Delta_{\underline{xy}}\right] \pi_1 \ . \tag{63a}$$

$$\rho_{\underline{xy}} = \sum_\alpha \pi_1 \exp[\ln R_{\underline{xy}}^\alpha + \Delta_{\underline{xy}}]\pi_1 \ , \tag{63b}$$

Eq.(63a) arises from combining the limits of Eqs.(62a) Eqs.(62b). Eq.(63b) arises from combining the limits of Eqs.(62c) and (62d). Note also that when all operators are diagonal and therefore commute, Eqs.(62c) and (62d) give Eq.(33), the definition of the classical $\Delta$. If we set $\pi_1 = 1$ and $\pi_0 = 0$ for now, then Eq.(63a) is the same as Eq.(57a). And Eq.(63b) is the same as Eq.(57b).

Now let us explain the purpose of the $\pi$ operators. Ideally, we would want to define $I^{(n+1)}$ by

$$I^{(n+1)} = \frac{1}{\sqrt{\rho_{\underline{xy}}}} \rho^{(n+1)} \frac{1}{\sqrt{\rho_{\underline{xy}}}} \ , \tag{64}$$

where

$$\rho^{(n+1)} = \sum_\alpha \tilde{K}_{\underline{xy}}^{\alpha(n+1)} \ . \tag{65}$$

However, if $\rho_{\underline{xy}}$ has any zero eigenvalues, its inverse square root does not exist. Note $\rho^{(n+1)}$ tends to $\rho_{\underline{xy}}$. For some small positive real $\epsilon$, we can define

$$\frac{1}{\sqrt{\rho_{\underline{xy}}}} \rho^{(n+1)} \frac{1}{\sqrt{\rho_{\underline{xy}}}} = \frac{1}{\sqrt{\rho_{\underline{xy}} + \epsilon\pi_0}} (\pi_1 \rho^{(n+1)} \pi_1 + \epsilon\pi_0) \frac{1}{\sqrt{\rho_{\underline{xy}} + \epsilon\pi_0}} \tag{66}$$

14

$$= \left( \pi_1 \frac{1}{\sqrt{\rho_{\underline{xy}}}} \pi_1 \right) \rho^{(n+1)} \left( \pi_1 \frac{1}{\sqrt{\rho_{\underline{xy}}}} \pi_1 \right) + \pi_0 \, . \tag{67}$$

(The $\epsilon\pi_0$ summands come into play only when $\rho_{\underline{xy}}$ vanishes.) This justifies Eq.(62c). $\rho_{\underline{xy}} = \sum_\alpha K^\alpha_{\underline{xy}}$ so we must also require that $K^\alpha_{\underline{xy}}$ vanish over the kernel space of $\rho_{\underline{xy}}$. We force this to happen by pre and post multiplying the right hand side of Eq.(63a) by $\pi_1$.

Another potential source of singular behavior in Eqs.(62) is the function $\exp(\ln R + D)$ where $R, D$ are Hermitian matrices and $R$ can be singular. This is not a theoretical disaster because even though the log of a zero eigenvalue of $R$ gives minus infinity, upon taking the exponential of that minus infinity, we get a zero contribution. From a numerical point of view, calculating $\exp(\ln R + D)$ accurately when $R$ is singular poses a challenge. Our first impulse is to calculate the eigenvalue expansion of $R$, take the log of the eigenvalues of the latter expansion, add $D$ to the result, calculate the eigenvalue expansion of $\ln R + D$, exponentiate the eigenvalues of the latter expansion. Finding the eigensystem of $\ln R + D$ can be hard to do accurately when $R$ is nearly singular and therefore $\ln R + D$ contains some nearly infinite eigenvalues. There are, however, other ways of exponentiating a matrix which do not require calculating its eigensystem. Ref.[15] describes 19 "dubious" ways of exponentiating a matrix. Its authors use the adjective dubious because none of these methods is ideal. Some work only for certain types of matrices, others entail an excessive number of operations, others are too sensitive, etc. In our case, whenever we use $\exp(\ln R + D)$, the matrix $\ln R + D$ is expected to be Hermitian with non-positive eigenvalues. Method 4 of Ref.[15] fits this situation perfectly. The method, first proposed by Colby et al in Ref.[16], is to approximate $\exp(-A)$ by a ratio of two $n$'th degree polynomials in A. The method works well even if some of the eigenvalues of $A$ are nearly infinite, as long as they are all non-negative.

## 5   Quantum Physics, Pure Minimization

In this section, we will discuss another quantum minimization problem. This minimization will differ from the quantum mixed minimization discussed previously in that now the range of our minimization will be restricted to those $\rho^\alpha_{\underline{xy}} \in \mathrm{dm}(\mathcal{H}_{\underline{xy}})$ of the special form:

$$\rho^\alpha_{\underline{xy}} = |\psi_\alpha\rangle\langle\psi_\alpha| \, , \tag{68}$$

where $|\psi_\alpha\rangle \in \mathcal{H}_{\underline{xy}}$.

As in the quantum mixed minimization problem, the CMI for $\rho_{\underline{xy}\alpha}$ can be expressed as

$$
\begin{aligned}
S_{\rho_{\underline{xy}\alpha}}(\underline{x}:\underline{y}|\underline{\alpha}) &= \sum_\alpha w_\alpha S_{\rho_{\underline{xy}}^\alpha}(\underline{x}:\underline{y}) \\
&= \sum_\alpha w_\alpha [S(\rho_{\underline{x}}^\alpha) + S(\rho_{\underline{y}}^\alpha) - S(\rho_{\underline{xy}}^\alpha)] \\
&= \sum_\alpha \mathrm{tr}_{\underline{x},\underline{y}}\left[K_{\underline{xy}}^\alpha(\log_2 K_{\underline{xy}}^\alpha - \log_2 R_{\underline{xy}}^\alpha)\right] \ .
\end{aligned}
\tag{69}
$$

Let

$$
\mathcal{K}_{pure}^\star = \{K_{\underline{xy}}^\bullet | \forall \alpha, \, K_{\underline{xy}}^\alpha = w_\alpha |\psi_\alpha\rangle\langle\psi_\alpha|, w_\bullet \in \mathrm{pd}(S_{\underline{\alpha}}), |\psi_\alpha\rangle \in \mathcal{H}_{\underline{xy}}\} \ ,
\tag{70}
$$

and

$$
\mathcal{K}_{pure} = \{K_{\underline{xy}}^\bullet \in \mathcal{K}_{pure}^\star | \sum_\alpha K_{\underline{xy}}^\alpha = \tilde{\rho}_{\underline{xy}}\} \ .
\tag{71}
$$

We define the entanglement $E_{pure}$ by

$$
E_{pure} = \left(\frac{1}{2}\right) \min_{K_{\underline{xy}}^\bullet \in \mathcal{K}_{pure}} S_{\rho_{\underline{xy}\alpha}}(\underline{x}:\underline{y}|\underline{\alpha}) \ .
\tag{72}
$$

In this definition of $E_{pure}$, $N_{\underline{\alpha}}$ will be assumed to tend to infinity. Ref.[14] shows that the limit is reached at a finite $N_{\underline{\alpha}} \le (N_{\underline{x}} N_{\underline{y}})^2$.

Note that since we are now assuming that the $\rho_{\underline{xy}}^\alpha$ are pure for all $\alpha$,

$$
\begin{aligned}
S_{\rho_{\underline{xy}}^\alpha}(\underline{x}:\underline{y}) &= S(\rho_{\underline{x}}^\alpha) + S(\rho_{\underline{y}}^\alpha) - S(\rho_{\underline{xy}}^\alpha) \\
&= 2S(\rho_{\underline{x}}^\alpha) \ .
\end{aligned}
\tag{73}
$$

Note also that

$$
\rho_{\underline{x}}^\alpha = \mathrm{tr}_{\underline{y}} |\psi_\alpha\rangle\langle\psi_\alpha| \ .
\tag{74}
$$

If $e = \{(w_\alpha, |\psi_\alpha\rangle)|\alpha \in S_{\underline{\alpha}}\}$ where $\rho_{\underline{xy}} = \sum_\alpha w_\alpha |\psi_\alpha\rangle\langle\psi_\alpha|$, we call $e$ a $\rho_{\underline{xy}}$ *ensemble or preparation*. Thus, our definition of $E_{pure}$ can be re-expressed as

$$
E_{pure} = \min_e \sum_\alpha w_\alpha S(\mathrm{tr}_{\underline{y}} |\psi_\alpha\rangle\langle\psi_\alpha|) \ ,
\tag{75}
$$

where the minimum is taken over all $\rho_{\underline{xy}}$ ensembles $e$. This is precisely the definition usually given for the entanglement of formation[18]. Thus, $E_{pure}$ is identical to the entanglement of formation.

It is convenient to consider the following "Lagrangian" functional of two density matrices:

$$\mathcal{L}(K^{\alpha}_{\underline{xy}}, K^{'\alpha}_{\underline{xy}}) = \sum_{\alpha} \mathrm{tr}_{\underline{x},\underline{y}} \left[ K^{\alpha}_{\underline{xy}} (\ln K^{\alpha}_{\underline{xy}} - \ln R^{'\alpha}_{\underline{xy}}) \right] \; , \tag{76}$$

where $R^{'\alpha}_{\underline{xy}} = R^{\alpha}_{\underline{xy}}[K^{'\alpha}_{\underline{xy}}]$. $E_{pure}$ can be defined in terms of this Lagrangian by

$$E_{pure} = \left( \frac{1}{2 \ln 2} \right) \min_{K^{\bullet}_{\underline{xy}} \in \mathcal{K}_{pure}} \mathcal{L}(K^{\alpha}_{\underline{xy}}, K^{\alpha}_{\underline{xy}}) \; . \tag{77}$$

**Lemma 5.1** $\mathcal{L}(K^{\alpha}_{\underline{xy}}, K^{'\alpha}_{\underline{xy}})$ *is convex ($\cup$) in its first argument.*

**proof:**

See proof of analogous lemma in the section on quantum mixed minimization.
QED

**Theorem 5.1** *Let*

$$\mathcal{K}'_{pure} = \mathcal{K}^{\star}_{pure} \; . \tag{78}$$

*At fixed $K^{\bullet}_{\underline{xy}} \in \mathcal{K}_{pure}$, $\mathcal{L}(K^{\alpha}_{\underline{xy}}, K^{'\alpha}_{\underline{xy}})$ is minimized over all $K^{'\bullet}_{\underline{xy}} \in \mathcal{K}'_{pure}$ iff $K^{'\alpha}_{\underline{xy}}$ satisfies*

$$w'_{\alpha} = w_{\alpha} \; , \tag{79a}$$

$$K^{'\alpha}_{\underline{x}} = K^{\alpha}_{\underline{x}} \; , \tag{79b}$$

*and*

$$K^{'\alpha}_{\underline{y}} = K^{\alpha}_{\underline{y}} \; . \tag{79c}$$

*Thus,*

$$\mathcal{L}(K^{\alpha}_{\underline{xy}}, K^{\alpha}_{\underline{xy}}) = \min_{K^{'\bullet}_{\underline{xy}} \in \mathcal{K}'_{pure}} \mathcal{L}(K^{\alpha}_{\underline{xy}}, K^{'\alpha}_{\underline{xy}}) \; , \tag{80}$$

*and*

$$E_{pure} = \left( \frac{1}{2 \ln 2} \right) \min_{K^{\bullet}_{\underline{xy}} \in \mathcal{K}_{pure}} \min_{K^{'\bullet}_{\underline{xy}} \in \mathcal{K}'_{pure}} \mathcal{L}(K^{\alpha}_{\underline{xy}}, K^{'\alpha}_{\underline{xy}}) \; . \tag{81}$$

**proof:**

See proof of analogous theorem in the section on quantum mixed minimization.
QED

**Theorem 5.2** *At fixed $K^{'\bullet}_{\underline{xy}} \in \mathcal{K}'_{pure}$, $\mathcal{L}(K^{\alpha}_{\underline{xy}}, K^{'\alpha}_{\underline{xy}})$ is minimized over all $K^{\bullet}_{\underline{xy}} \in \mathcal{K}_{pure}$ iff $K^{\alpha}_{\underline{xy}} = w_{\alpha}|\psi_{\alpha}\rangle\langle\psi_{\alpha}|$ satisfies*

$$\ln(w_{\alpha})|\psi_{\alpha}\rangle = (\ln R^{'\alpha}_{\underline{xy}} + \Delta_{\underline{xy}})|\psi_{\alpha}\rangle \; , \tag{82a}$$

17

*and*

$$\rho_{\underline{xy}} = \sum_\alpha w_\alpha |\psi_\alpha\rangle\langle\psi_\alpha| \ . \tag{82b}$$

**proof:**

Let

$$|n_\alpha\rangle = \sqrt{w_\alpha}|\psi_\alpha\rangle \tag{83}$$

and

$$A_\alpha = \ln K_{\underline{xy}}^\alpha - \ln R_{\underline{xy}}^{'\alpha} \ . \tag{84}$$

Then the Lagrangian $\mathcal{L}$ can be expressed as:

$$\mathcal{L}(K_{\underline{xy}}^\alpha, K_{\underline{xy}}^{'\alpha}) = \sum_\alpha \mathrm{tr}_{\underline{xy}}\left(|n_\alpha\rangle\langle n_\alpha|A_\alpha\right) \ . \tag{85}$$

Suppose a minimum is achieved. Define a new Lagrangian $\mathcal{L}_{tot}$ by adding to $\mathcal{L}$ a Lagrange multiplier term that enforces the constraint that the sum over $\alpha$ of $|n_\alpha\rangle\langle n_\alpha|$ equals a fixed density matrix $\tilde{\rho}_{\underline{xy}} \in \mathrm{dm}(\mathcal{H}_{\underline{xy}})$:

$$\mathcal{L}_{tot} = \mathcal{L}(K_{\underline{xy}}^\alpha, K_{\underline{xy}}^{'\alpha}) + \mathrm{tr}_{\underline{x},\underline{y}}[\lambda_{\underline{xy}}(\sum_\alpha |n_\alpha\rangle\langle n_\alpha| - \tilde{\rho}_{\underline{xy}})] \ . \tag{86}$$

$\mathcal{L}_{tot}$ should not change if we vary infinitesimally and independently the operators $\lambda_{\underline{xy}}$ and $|n_\alpha\rangle\langle n_\alpha|$ for all $\alpha$. Thus,

$$0 = \delta\mathcal{L}_{tot} = \sum_\alpha \mathrm{tr}_{\underline{x},\underline{y}}\left[\delta(|n_\alpha\rangle\langle n_\alpha|)(A_\alpha + 1 + \lambda_{\underline{xy}})\right] \ . \tag{87}$$

Suppose we have an arbitrary Hermitian operator $A$ acting on some Hilbert space $\mathcal{H}$ and $|n\rangle \in \mathcal{H}$. If

$$0 = \mathrm{tr}\left[\delta(|n\rangle\langle n|)A\right] \ , \tag{88}$$

then

$$0 = \langle n|A\delta(|n\rangle) + (\delta\langle n|)A|n\rangle \tag{89}$$

so

$$A|n\rangle = 0 \ . \tag{90}$$

Let

$$\Delta_{\underline{xy}} = -1 - \lambda_{\underline{xy}} \ . \tag{91}$$

Then Eq.(87) implies that

18

$$(\ln K_{\underline{xy}}^\alpha - \ln R_{\underline{xy}}^{'\alpha} - \Delta_{\underline{xy}})|\psi_\alpha\rangle = 0 \ . \tag{92}$$

(Assume $w_\alpha \neq 0$ for all $\alpha$). Suppose $w \in (0, 1]$. Given any column vector $|\psi\rangle \in \mathcal{H}$, we can always find a unitary matrix $U$ such that $|\psi\rangle = U|0\rangle$, where $|0\rangle$ is the unit vector which has one as its first component and zero for all others. Thus

$$
\begin{aligned}
\ln\left(w|\psi\rangle\langle\psi|\right)|\psi\rangle &= U\ln\left(w|0\rangle\langle 0|\right)U^\dagger U|0\rangle \\
&= U\,diag(\ln w, -\infty, -\infty, \ldots)\begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \end{bmatrix} \\
&= \ln(w)|\psi\rangle \ .
\end{aligned}
\tag{93}
$$

Thus, Eq.(92) implies that

$$\ln w_\alpha|\psi_\alpha\rangle = (\ln R_{\underline{xy}}^{'\alpha} + \Delta_{\underline{xy}})|\psi_\alpha\rangle \ . \tag{94}$$

$\mathcal{L}(K, K')$ has an extremum in $K$ at fixed $K'$ iff Eqs.(82) are true. Furthermore, since $\mathcal{L}$ is convex in its first argument, the extremum must be a global minimum. QED

**Theorem 5.3** *The CMI minimum which defines $E_{pure}$ is achieved iff*

$$\ln(w_\alpha)|\psi_\alpha\rangle = (\ln R_{\underline{xy}}^\alpha + \Delta_{\underline{xy}})|\psi_\alpha\rangle \ , \tag{95a}$$

*and*

$$\rho_{\underline{xy}} = \sum_\alpha w_\alpha|\psi_\alpha\rangle\langle\psi_\alpha| \ . \tag{95b}$$

*Furthermore,*

$$E_{pure} = \left(\frac{1}{2\ln 2}\right)\langle\Delta\rangle \ , \tag{96}$$

*where*

$$\langle\Delta\rangle = \mathrm{tr}_{\underline{x},\underline{y}}(\rho_{\underline{xy}}\Delta_{\underline{xy}}) \ . \tag{97}$$

*This justifies calling $\Delta_{\underline{xy}}$ an "entanglement operator".*

**proof:**

The first part of this claim just brings together results obtained in the previous two theorems. For a proof of the second part where $E_{pure}$ is expressed in terms of $\Delta$, see the analogous theorem for quantum mixed minimization. QED

As in the quantum mixed minimization problem, we will define two sequences of operators, $K_{\underline{xy}}^{\alpha(n)} = w_\alpha^{(n)}|\psi_\alpha^{(n)}\rangle\langle\psi_\alpha^{(n)}|$ and $\Delta_{\underline{xy}}^{(n)}$ for $n = 0, 1, 2 \ldots$. The sequences will

19

converge to $K_{\underline{xy}}^{\alpha}$ and $\Delta_{\underline{xy}}$, respectively, as $n \to \infty$. We will define our two sequences recursively. Let $K_{\underline{xy}}^{\alpha(0)}$ be chosen arbitrarily from $\mathcal{K}_{pure}$. Let $\Delta_{\underline{xy}}^{(0)} = 0$. For any $n \geq 0$, let

$$w_{\alpha}^{(n+1)}|\psi_{\alpha}^{(n+1)}\rangle = \pi_1 \exp\left[\ln R_{\underline{xy}}^{\alpha(n)} + \Delta_{\underline{xy}}^{(n)}\right] \pi_1 |\psi_{\alpha}^{(n+1)}\rangle , \tag{98a}$$

$$K_{\underline{xy}}^{\alpha(n+1)} = \frac{w_{\alpha}^{(n+1)}|\psi_{\alpha}^{(n+1)}\rangle\langle\psi_{\alpha}^{(n+1)}|}{\sum_{\alpha} \mathrm{tr}_{\underline{xy}}(numerator)} , \tag{98b}$$

$$\tilde{w}_{\alpha}^{(n+1)}|\tilde{\psi}_{\alpha}^{(n+1)}\rangle = \pi_1 \exp\left[\ln R_{\underline{xy}}^{\alpha(n+1)} + \Delta_{\underline{xy}}^{(n)}\right] \pi_1 |\tilde{\psi}_{\alpha}^{(n+1)}\rangle , \tag{98c}$$

$$\tilde{K}_{\underline{xy}}^{\alpha(n+1)} = \frac{\tilde{w}_{\alpha}^{(n+1)}|\tilde{\psi}_{\alpha}^{(n+1)}\rangle\langle\tilde{\psi}_{\alpha}^{(n+1)}|}{\sum_{\alpha} \mathrm{tr}_{\underline{xy}}(numerator)} , \tag{98d}$$

$$I^{(n+1)} = \left(\pi_1 \frac{1}{\sqrt{\rho_{\underline{xy}}}}\pi_1\right)\left\{\sum_{\alpha} \tilde{K}_{\underline{xy}}^{\alpha(n+1)}\right\}\left(\pi_1 \frac{1}{\sqrt{\rho_{\underline{xy}}}}\pi_1\right) + \pi_0 , \tag{98e}$$

$$\Delta_{\underline{xy}}^{(n+1)} = -\ln\left(e^{-\frac{\Delta_{\underline{xy}}^{(n)}}{2}} I^{(n+1)} e^{-\frac{\Delta_{\underline{xy}}^{(n)}}{2}}\right) , \tag{98f}$$

where $R_{\underline{xy}}^{\alpha(n)} = R_{\underline{xy}}^{\alpha}[K_{\underline{xy}}^{\alpha(n)}]$, $\pi_0 = \pi_{ker}(\rho_{\underline{xy}})$ and $\pi_1 = 1 - \pi_0$. In Eqs.(98a) and (98c), we choose the **largest** eigenvalue and the corresponding eigenvector of the operator on the right hand side of the equation. As $n$ tends to infinity, said operator tends towards a projection operator, so its eigenvalues all go to zero except for possibly one of them.

In this paper, we won't prove that the sequences $K_{\underline{xy}}^{\alpha(n)}$ and $\Delta_{\underline{xy}}^{(n)}$ converge. We defer that to future papers, confining ourselves here to presenting some empirical and intuitive motivations for the sequences. In Section 6, we give some computer results that are good evidence of convergence. It is easy to see that if the sequences do converge, then their limit satisfies Eqs.(95).

# 6 Causa Común

We have written a C++ program called Causa Común ("Common Cause" in Spanish). It's called this because entanglement is a manifestation of causality; it occurs between several events with a common cause. Causa Común can do all three minimizations considered in this paper (classical, quantum mixed and quantum pure). For each of these three cases, it can find the entanglement, entanglement operator, and an optimal state decomposition. Next, we will discuss Causa Común output for two examples of quantum states: Bell Mixtures, and Horodecki States.

## 6.1 Bell Mixtures

In this example

$$S_{\underline{x}} = S_{\underline{y}} = Bool \ . \qquad (99)$$

The following four states are usually called the "Bell basis" of $\mathcal{H}_{\underline{xy}}$:

$$|\psi^{\pm}\rangle = |\neq^{\pm}\rangle = \frac{1}{\sqrt{2}}(|01\rangle \pm |10\rangle) \ , \qquad (100)$$

and

$$|\phi^{\pm}\rangle = |=^{\pm}\rangle = \frac{1}{\sqrt{2}}(|00\rangle \pm |11\rangle) \ . \qquad (101)$$

Let $|B(\mu)\rangle$ with $\mu \in \{0, 1, 2, 3\}$ represent the four Bell states. Call a "Bell mixture" any density matrix $\rho_{\underline{xy}}$ expressible in the form

$$\rho_{\underline{xy}} = \sum_{\mu} m_{\mu}|B(\mu)\rangle\langle B(\mu)| \ , \qquad (102)$$

where $m_{\bullet} \in \mathrm{pd}(S_{\underline{\mu}})$. For any $p \in [0, 1]$, define the *binary entropy* $h(p)$ by

$$h(p) = -[p \log_2(p) + (1 - p) \log_2(1 - p)] \ . \qquad (103)$$

Ref.[18] showed that for any Bell mixture $\rho_{\underline{xy}}$, the entanglement of formation is given by:

$$E_{form.}(\rho_{\underline{xy}}) = h\left(\frac{1 + \sqrt{1 - t}}{2}\right) \ , \qquad (104a)$$

$$t = \begin{cases} 0 & \text{if } m_{max} < \frac{1}{2} \\ (2m_{max} - 1)^2 & \text{otherwise} \end{cases} \ , \qquad (104b)$$

where $m_{max}$ refers to the maximum of the weights $m_{\mu}$.

We will refer to the following one parameter family of Bell mixtures as the "Werner States":

$$W(F) = F|=^{+}\rangle\langle=^{+}| + \frac{(1 - F)}{3}\left(|=^{-}\rangle\langle=^{-}| + |\neq^{+}\rangle\langle\neq^{+}| + |\neq^{-}\rangle\langle\neq^{-}|\right) \ , \qquad (105)$$

for $F \in [0, 1]$.

Fig.1 is Causa Común output for $E_{pure}$ and $E_{mixed}$ of the Werner States using $N_{\underline{\alpha}} = 6$. We also got $E_{mixed} = E_{pure} = 0$ for $0 \leq F \leq \frac{1}{2}$ (not shown in graph). We compared $E_{pure}$ obtained using our algorithm and $E_{form}$ obtained using Eq.(104), and found them to agree well over the entire range $F \in [0, 1]$. Note that $E_{mixed}$ behaves like the entanglement of distillation: both are non-negative, less than or equal to $E_{form}$, and equal to $E_{form}$ for pure density matrices. In a future paper, we will explain the close connection between $E_{mixed}$ and entanglement of distillation.
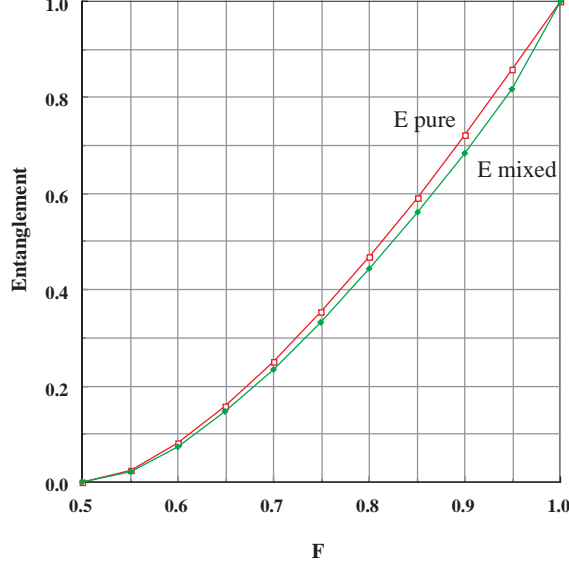
Figure 1: Pure and mixed entanglements for Werner States.

## 6.2 Horodecki States

In this example,

$$S_{\underline{x}} = S_{\underline{y}} = \{0, 1, 2\} . \tag{106}$$

Let

$$\langle x, y | \psi^+ \rangle = \frac{1}{\sqrt{3}} (\delta_{xy}^{00} + \delta_{xy}^{11} + \delta_{xy}^{22}) , \tag{107}$$

$$(\sigma_{\underline{xy}}^+)_{xy,x'y'} = \frac{\delta_{xy}^{x'y'}}{3} (\delta_{xy}^{01} + \delta_{xy}^{12} + \delta_{xy}^{20}) , \tag{108}$$

$$(\sigma_{\underline{xy}}^-)_{xy,x'y'} = \frac{\delta_{xy}^{x'y'}}{3} (\delta_{xy}^{10} + \delta_{xy}^{21} + \delta_{xy}^{02}) . \tag{109}$$

We will refer to the following one parameter family of density matrices as the "Horodecki States":

$$\sigma(\alpha) = \frac{2}{7} |\psi^+\rangle\langle\psi^+| + \frac{\alpha}{7}\sigma^+ + \frac{5-\alpha}{7}\sigma^- , \tag{110}$$

where $\alpha \in [2, 5]$. These states where first introduced in Ref.([19]), where it was shown that they are separable for $\alpha \in (2, 3)$, bound entangled for $\alpha \in (3, 4)$, and free entangled for $\alpha \in (4, 5)$.
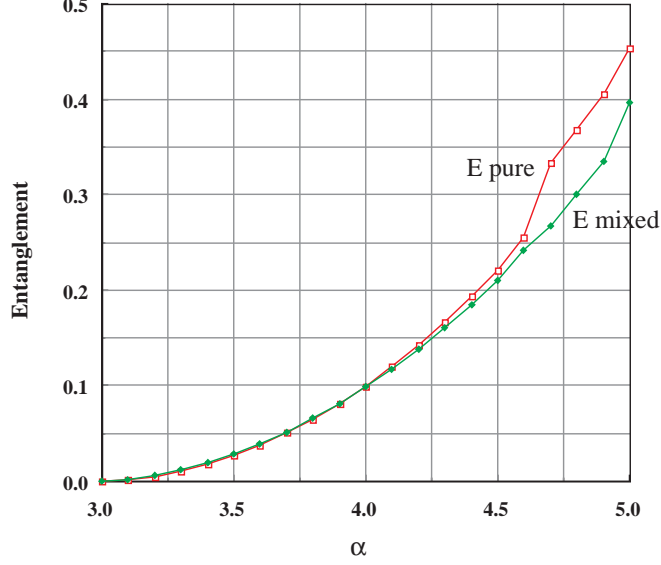
22

Figure 2: Pure and mixed entanglements for Horodecki States.

Fig.2 is Causa Común output for $E_{pure}$ and $E_{mixed}$ of the Horodecki States using $N_{\underline{\alpha}} = 12$. We also got $E_{mixed} = E_{pure} = 0$ for $2 \leq \alpha \leq 3$ (not shown in graph).[20]

# Acknowledgements

# References

[1] R.R. Tucci, "Quantum Entanglement and Conditional Information Transmission", Los Alamos eprint quant-ph/9909040 .

[2] R.R. Tucci, "Separability of Density Matrices and Conditional Information Transmission", Los Alamos eprint quant-ph/0005119 .

[3] R.R. Tucci, "Entanglement of formation and conditional information transmission", Los Alamos eprint quant-ph/0010041 .

[4] S. Arimoto, "An algorithm for calculating the capacity of an arbitrary discrete memoryless channel", IEEE Trans. Info Theory **18**, 14 (1972).

[5] R. E. Blahut, "Computation of channel capacity and rate distortion functions", IEEE Trans. Info Theory **18**, 460 (1972).

[6] R.E. Blahut, "Principles and Practice of Information Theory", (Addison-Wesley, 1987) pgs 164-170.

[7] S. Osawa, H. Nagaoka, "Numerical experiments on the capacity of a quantum channel with entangled inputs", Los Alamos eprint quant-ph/0007115 .

[8] K. Zyczkowski, "Volume of the set of entangled states II", Los Alamos eprint quant-ph/9902050 .

[9] K. Audenaert, F. Verstraete, B. De Moor, "Variational characterizations of separability and entanglement of formation", Los Alamos eprint quant-ph/0006128 .

[10] R.R. Tucci, "Quantum Information Theory - A Quantum Bayesian Nets Perspective", Los Alamos eprint quant-ph/9909039 .

[11] Thomas M. Cover and Joy A. Thomas, "Elements of Information Theory (2cnd edition)" (Wiley, 1991).

[12] A. Wehrl, "General Properties of Entropy", Rev. Mod. Phys. **50** 221-260 (1978).

[13] Call diagram Eq.(61) a K-jumps-first strategy. It is also possible to use a $\Delta$-jumps-first strategy defined by the following diagram:

$$
\begin{array}{ccccccc}
K_{\underline{xy}}^{\alpha(0)} & \to & K_{\underline{xy}}^{\alpha(1)} & \to & K_{\underline{xy}}^{\alpha(2)} & \cdots & \\
& \searrow & \uparrow & \searrow & \uparrow & & \\
\Delta_{\underline{xy}}^{(0)} & \to & \Delta_{\underline{xy}}^{(1)} & \to & \Delta_{\underline{xy}}^{(2)} & \cdots &
\end{array}
\tag{111}
$$

We won't discuss the $\Delta$-jumps-first strategy in this paper because it is very similar to the $K$-jumps-first one. (The $\Delta$-jumps-first strategy appears to require more numerical operations to move from $n$ to $n+1$)

[14] P. Horodecki, "Separability Criterion and Inseparable Mixed States with Positive Partial Transpose", Los Alamos eprint quant-ph/9703004 .

[15] C.B. Moller, C.F. Van Loan, "19 dubious ways to compute the exponential of a matrix", SIAM Review **20**, 801 (1979).

[16] W.J. Cody, G. Meinardus, R.S. Varga, "Chebyshev Rational Approximations to $e^{-x}$ in $[0, +\infty)$ and applications to heat conduction problems", Jour. of Approx. Theo. **2**, 50 (1969).

[17] L.P. Hughston, R. Jozsa, W.K. Wootters, "A complete classification of quantum ensembles having a given density matrix", Phys. Let. A **183** (1993) 14.

[18] C.H. Bennett, D.P. DiVincenzo, J. Smolin, W.K. Wootters, "Mixed state entanglement and quantum error correction", Los Alamos eprint quant-ph/9604024 .

[19] P. Horodecki, "Bound entanglement can be activated", Los Alamos eprint quant-ph/9806058 .

[20] K. Audenaert (KA) et al have written a computer program that calculates entanglement of formation using a method that is very different from mine. KA kindly provided me with a plot generated by his program of $E_{form}$ versus $\alpha$ for the Horodecki States. I compared his $E_{form}$ with my $E_{pure}$. I found both graphs to agree very well for $\alpha \in [2, 4.5]$ and for $\alpha = 5$. His graph connects the points at $\alpha = 4.5$ and $\alpha = 5$ in a smooth way. Mine is higher in that range although we agree at 4.5 and 5. I believe his plot to be the more accurate. I believe the source of our disagreement is that Causa Común is getting stuck in non-global minima.

KA et al reported in Ref.([9]) that they routinely run their program with about ten starting points and then choose the smallest $E_{form}$ of the ten. They do this because they have found multiple minima in the function being minimized. I too average over several (five) starting points as I too have found multiple minima.

I am currently using poorly motivated starting points: I generate a random right unitary matrix $T_j^\alpha$ and construct $K_{\underline{xy}}^{\alpha(0)}$ from $T_j^\alpha$ and $\rho_{\underline{xy}}$. KA et al, on the other hand, use a technique (see Ref.([9]) which gives them a more refined guess as their starting point. I intend to incorporate a similar technique into Causa Común in the future. I believe that by choosing a starting point $K_{\underline{xy}}^{\alpha(0)}$ with the appropriate symmetry, one can significantly improve the chances that the algorithm will converge to the global minimum. An analogous issue arises when using a variational method to solve for the quantum stationary states of a particle in a potential well. Using a trial wavefunction with the appropriate symmetry (e.g., no nodes) helps the algorithm converge to the ground state energy level.